Risk factors for Salmon Rickettsial Septicaemia (SRS) and Caligus infestation in farmed salmonid fish in Chile: design of a retrospective study

Ausvet Pty Ltd

PROGRAMA
PARA LA GESTIÓN SANITARIA EN LA
ACUICULTURA

TABLE OF CONTENTS

INTRODUCTION

Salmon Rickettsial Septicaemia and Caligus infestation are two of the most economically important diseases of salmonid fish farmed in marine waters of Chile. The effective and efficient control of these disease depends on a good understanding of the factors that influence transmission of infection, and the impact and duration of outbreaks once transmission has occurred.

This section of the report outlines a retrospective study to enhance knowledge about risk and protective factors of key diseases of the Chilean salmonid fish industry. An important opportunity for this study has arisen through the recent development of an integrated data management system that is expected to enable sophisticated analysis of a very large dataset. The study will assess a broad range of factors for which data is available, focussing on management factors that could be adjusted to decrease the impact of disease and the costs of production.

The study is expected to provide valuable insights and learnings that will inform the design of future (prospective) studies and research.

BACKGROUND

In the past two decades, substantial efforts have been made across the world to increase knowledge about SRS and various species of sea louse. Research projects and analyses of various government and industry data sources have generated a body of scientific knowledge and hypotheses about the patterns of these diseases. However, key questions about the epidemiology of the diseases remain unanswered and the ongoing impacts and production costs associated remain very high.

While there have been several very good analyses in recent years, others have been limited by the quality and quantity of data available to the researchers. For example, some reports have presented crude associations but have been unable to fully account for possible confounders and interactions. Other studies have been relative small and lacked power to detect differences in disease outcomes between groups with different exposures to potential risk factors or protective factors.

Retrospective analyses of data are often constrained by the quality and quantity of available data because research questions, study designs and data collection protocols were not established before the data were collected. However, the recent integration of various data sources in a single database and the subsequent availability of a very large dataset may allow a level of analysis that has not previously been possible.

AIMS AND OBJECTIVES

The study aims to add to the existing body of knowledge about the distribution of important diseases of salmonid fish in Chile and factors that influence the incidence and severity of these diseases.

The primary objective is to answer several key questions:

1. **Where, when and to what extent do SRS and Caligus affect salmonid fish farmed in Chile?**

   This part of the study aims to describe the frequency (incidence or prevalence) and spatial distribution of these diseases in Atlantic salmon, Coho salmon and rainbow trout during the period for which data is available.

2. **What biological, environmental and management factors appear to influence the incidence and severity of these diseases?**

   This part of the study aims to describe the distribution of possible risk factors (including protective factors) and identify any statistically significant associations between these factors and disease occurrence.

3. **What hypotheses can be generated about the causes of these disease and how could these hypotheses be tested?**

   Studies of this type cannot provide proof of causal associations, but may provide insights that enable us to generation of hypotheses about why and how disease occurs in specific populations at specific times.

The **secondary objectives** of the study are:

- to identify constraints and opportunities—for example, relating to data collection, data management and analytical methods—that can be addressed in prospective studies.
- to build capacity for analysis and reporting—skills and knowledge, as well as functionality of the new integrated management system—that will have sustainable benefits.

STUDY DESIGN

The study is a retrospective observational study based on an analysis of data compiled from multiple sources over recent years. We propose a cross-sectional study that explores associations between selected exposures (risk factors) and outcomes (for example, disease events and production indices) at the cage, farm and/or barrio level during a defined period of the production cycle.

Several a priori hypotheses about causal associations (between specific risk factors and disease outcomes) will be developed through consultation with stakeholders before embarking on detailed analysis: these will become focuses of the analyses.

Depending on the scope and quality of available data, it may be possible to consider other study designs including 'historical' cohort studies and case-control studies. This will be considered during refinement of the study design after initial exploration of the available data.

The scope of the retrospective data will be dependent on the data made available through the Omnísyan data management framework.

DEFINING EXPOSURES AND OUTCOMES

It is anticipated that the analyses will use a variety of data on exposures (for example, management practices, vaccinations/treatments and environmental factors) and outcomes (for example, production indicators and mortality). At an early stage of the study, it will be necessary to define the exposures (including interventions) and outcomes that will be the focus of the study. This is important for several reasons, including:

1. To reduce misclassification of exposure status and outcome status. For example, a clear definition of an 'outbreak' is required if we want to analyze time-to-event data such as time to first outbreak of SRS after different vaccination protocols.

2. To enable analysts to select the most appropriate methods of analysis. For example, is it better to define Caligus infestation on a continuous scale (in which case multivariable linear regression would be appropriate) or as a binary variable—e.g. above or below a treatment threshold—in which case multivariable logistic regression might be best?

3. To identify the factors that are most relevant from a production perspective. Environmental factors such as water temperature may influence disease occurrence, but we acknowledge the importance of focusing on those management and genetic factors that could be changed.

The approach to analysing the data will depend on the data available and the ways in which key exposures and outcomes are defined. It will be useful to hold discussions with epidemiologists and experts in salmonid fish production to identify and define these exposures and outcomes.

DATA AND DATA MANAGEMENT

This study will take advantage of a very large dataset—including data on possible risk factors and variety of outcome measures—that has been integrated in the Omnísyan framework. This framework greatly enhances capacity to access and analyse data from sources including the SIFA and INFA databases.

APPROACH TO THE ANALYSIS

The approach to the analysis will depend on the nature of the data available and the research questions that the analysis seeks to address. The following is a guide, based on the objectives described above, but it will be refined after further discussion with collaborators and initial exploration of the data.

Initial exploration of the data will inform selection of appropriate study types (longitudinal, historical cohort-study, case-control) and methods of analysis. For example, some analyses will depend on whether data is available to accurately determine time-at-risk.

To address Question 1 (distribution of disease), descriptive techniques will be used to generate a comprehensive picture of disease occurrence in time and space over the duration of the period for which data are available. It will present information about the population at risk during this period (denominator data) and identify trends and patterns of disease outcomes in time and space. This part of the analysis will take advantage of the tools for simple epidemiological analysis and data visualisation developed within the Omnísyan framework.

Question 2 (identification of risk factors) is an essential component of the study and several analytical methods may be suitable for this kind of data. Initial analysis will estimate crude

measures of association between possible risk factors and disease outcomes. More advanced statistical methods (such as multivariable regression modelling or survival analysis) and spatial epidemiological analyses will then be applied to identify and estimate the effect of key risk factors, taking into account interactions and adjusting for possible confounding factors. Advanced spatial analyses may allow detection of spatiotemporal clusters. Information theoretic approaches—which build on existing knowledge and a priori hypotheses about disease causality—may offer an alternative to traditional model-building approaches.

Because this study is based on existing data, it is not necessary to conduct sample size and power calculations at the beginning. However, post-hoc estimates of the power of the study will be conducted to show the Type II error rate. For example, if the results suggest that a risk factor is not associated with a disease outcome, power calculations can estimate the probability that a real difference was not detected.

The results of the analysis will be shared with stakeholders to encourage discussion about possible causes and generate hypotheses that could be tested through future studies.

INTERPRETATION OF RESULTS

Studies of this type are efficient and useful for generating hypotheses about potential causal relationship and preliminary testing of a priori hypotheses. The results of analyses will be interpreted with consideration to the origin of the data and possible sources of bias. If the available data are of reasonable completeness and quality, it may be possible to suggest which risk factors and protective factors are likely to have the greatest association with the occurrence of disease and production-related outcomes; however, limitations to historical data often make it difficult to infer causality from apparent associations. In addition, confounding may be imperfectly controlled and estimates derived from the analysis may be biased. For example, if outcomes are measured as prevalence, it can be difficult to determine if a factor is associated with an outcome because it increased the persistence of the outcome (as might happen with genetic characteristics that increase resistance to disease in the face of infection pressure) or if it is associated with the development of the outcome in the first place.

Finally, if the exposures (such as management practices) change over time, it can be difficult to differentiate cause and effect (the 'reverse causation problem'); for example, if a certain management practice appeared to be positively associated with the prevalence of Caligus infestation, it may be difficult to differentiate between farms that initiated the management practice in response to a problem with Caligus and those farms where the infestation occurred because of the management practice.

If multiple hypotheses about associations are tested in this study, then there is a risk of Type 1 errors (i.e. concluding that a real effect is present when in fact there is none). However, information theoretic approaches may be used to minimise Type 1 errors.

SUSTAINABLE BENEFITS

This study is expected to provide sustainable benefits in the context of a larger program of work to strengthen the control of infectious diseases of salmonid fish in Chile. For example:

- This study will be conducted early in the program, so lessons learned can inform the design and implementation of prospective studies. Any important gaps or problems in existing data can be identified and corrected in future.
- Advanced analysis of historical data will enable the development of data management tools and replicable analyses/reports that may be used again for future studies.
- Collaboration between Sernapesca, consultants and other researchers may offer opportunities for training and professional development.

RESOURCES

The retrospective study will be led by consultants with expertise in epidemiology, biostatistics and visualisation/communication of scientific information. However, it would be preferable to conduct the study as a partnership, in close collaboration with the industry representatives, researchers and policy-makers that are the primary users of the study results.

It is important to ensure that the necessary data is available for analysis.

COMMUNICATION

This study will be conducted to address important concerns of aquaculture industries and government agencies in Chile. If it is to be effective, it will be important to communicate well during refinement of the design and during implementation. It would be useful to identify the target audiences for the outputs of this study and discuss ways to invite design inputs and disseminate results.


PROGRAMA
PARA LA GESTIÓN SANITARIA EN LA
ACUICULTURA